

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Garoff, Henrik
Liljestrom, Peter
- (ii) TITLE OF INVENTION: DNA Expression Systems Based on
Alphaviruses

(iii) NUMBER OF SEQUENCES: 27

BB
DO NOT TYPE OR FAX
BB

- (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Birch, Stewart, Kolasch & Birch
- (B) STREET: P.O. Box 747
- (C) CITY: Falls Church
- (D) STATE: Virginia
- (E) COUNTRY: USA
- (F) ZIP: 22040-0747

- (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

- (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: US 07/920,281
- (B) FILING DATE: 13-AUG-1992
- (C) CLASSIFICATION:

- (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Murphy Jr., Gerald M.
- (B) REGISTRATION NUMBER: 28,977
- (C) REFERENCE/DOCKET NUMBER: 828-103P

- (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 703-241-1300
- (B) TELEFAX: 703-241-2848
- (C) TELEX: 248345

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11517 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Semliki Forest Virus

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..11517

(D) OTHER INFORMATION: /label= genome
 /note= "Semliki Forest Virus complete nucleotide sequence, presented as a cloned DNA sequence; see Figure 5."

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 87..7379

(D) OTHER INFORMATION: /product= "SFV polyprotein"

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 7421..11179

(D) OTHER INFORMATION: /product= "SFV polyprotein"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GATGGCGGAT GTGTGACATA CACGACGCCA AAAGATTTG TTCCAGCTCC TGCCACCTCC	60
GCTACGCGAG AGATTAACCA CCCACG ATG GCC GCC AAA GTG CAT GTT GAT ATT	113
Met Ala Ala Lys Val His Val Asp Ile	
1 5	
GAG GCT GAC AGC CCA TTC ATC AAG TCT TTG CAG AAG GCA TTT CCG TCG	161
Glu Ala Asp Ser Pro Phe Ile Lys Ser Leu Gln Lys Ala Phe Pro Ser	
10 15 20 25	
TTC GAG GTG GAG TCA TTG CAG GTC ACA CCA AAT GAC CAT GCA AAT GCC	209
Phe Glu Val Glu Ser Leu Gln Val Thr Pro Asn Asp His Ala Asn Ala	
30 35 40	
AGA GCA TTT TCG CAC CTG GCT ACC AAA TTG ATC GAG CAG GAG ACT GAC	257
Arg Ala Phe Ser His Leu Ala Thr Lys Leu Ile Glu Gln Glu Thr Asp	
45 50 55	
AAA GAC ACA CTC ATC TTG GAT ATC GGC AGT GCG CCT TCC AGG AGA ATG	305
Lys Asp Thr Leu Ile Leu Asp Ile Gly Ser Ala Pro Ser Arg Arg Met	
60 65 70	

ATG TCT ACG CAC AAA TAC CAC TGC GTC CCT ATG CGC AGC GCA GAA Met Ser Thr His Lys Tyr His Cys Val Cys Pro Met Arg Ser Ala Glu	75	80	85	353
GAC CCC GAA AGG CTC GAT AGC TAC GCA AAG AAA CTG GCA GCG GCC TCC Asp Pro Glu Arg Leu Asp Ser Tyr Ala Lys Lys Leu Ala Ala Ala Ser	90	95	100	401
GGG AAG GTG CTG GAT AGA GAG ATC GCA GGA AAA ATC ACC GAC CTG CAG Gly Lys Val Leu Asp Arg Glu Ile Ala Gly Lys Ile Thr Asp Leu Gln	110	115	120	449
ACC GTC ATG GCT ACG CCA GAC GCT GAA TCT CCT ACC TTT TGC CTG CAT Thr Val Met Ala Thr Pro Asp Ala Glu Ser Pro Thr Phe Cys Leu His	125	130	135	497
ACA GAC GTC ACG TGT CGT ACG GCA GCC GAA GTG GCC GTA TAC CAG GAC Thr Asp Val Thr Cys Arg Thr Ala Ala Glu Val Ala Val Tyr Gln Asp	140	145	150	545
GTG TAT GCT GTA CAT GCA CCA ACA TCG CTG TAC CAT CAG GCG ATG AAA Val Tyr Ala Val His Ala Pro Thr Ser Leu Tyr His Gln Ala Met Lys	155	160	165	593
TGT GTC AGA ACG GCG TAT TGG ATT GGG TTT GAC ACC ACC CCG TTT ATG Gly Val Arg Thr Ala Tyr Trp Ile Gly Phe Asp Thr Thr Pro Phe Met	170	175	180	641
TTC GAC GCG CTA GCA GGC GCG TAT CCA ACC TAC GCC ACA AAC TGG GCC Phe Asp Ala Leu Ala Gly Ala Tyr Pro Thr Tyr Ala Thr Asn Trp Ala	190	195	200	689
GAC GAG CAG GTG TTA CAG GCC AGG AAC ATA GGA CTG TGT GCA GCA TCC Asp Glu Gln Val Leu Gln Ala Arg Asn Ile Gly Leu Cys Ala Ala Ser	205	210	215	737
TTG ACT GAG GGA AGA CTC GGC AAA CTG TCC ATT CTC CGC AAG AAG CAA Leu Thr Glu Gly Arg Leu Gly Lys Leu Ser Ile Leu Arg Lys Lys Gln	220	225	230	785
TTG AAA CCT TGC GAC ACA GTC ATG TTC TCG GTA GGA TCT ACA TTG TAC Leu Lys Pro Cys Asp Thr Val Met Phe Ser Val Gly Ser Thr Leu Tyr	235	240	245	833
ACT GAG AGC AGA AAG CTA CTG AGG AGC TGG CAC TTA CCC TCC GTA TTC Thr Glu Ser Arg Lys Leu Leu Arg Ser Trp His Leu Pro Ser Val Phe	250	255	260	881
CAC CTG AAA GGT AAA CAA TCC TTT ACC TGT AGG TGC GAT ACC ATC GTA His Leu Lys Gly Lys Gln Ser Phe Thr Cys Arg Cys Asp Thr Ile Val	270	275	280	929

TCA TGT GAA GGG TAC GTA GTT AAG AAA ATC ACT ATG TGC CCC GGC CTG Ser Cys Glu Gly Tyr Val Val Lys Lys Ile Thr Met Cys Pro Gly Leu 285 290 295	977
TAC GGT AAA ACG GTA GGG TAC GCC GTG ACG TAT CAC GCG GAG GGA TTC Tyr Gly Lys Thr Val Gly Tyr Ala Val Thr Tyr His Ala Glu Gly Phe 300 305 310	1025
CTA GTG TGC AAG ACC ACA GAC ACT GTC AAA GGA GAA AGA GTC TCA TTC Leu Val Cys Lys Thr Thr Asp Thr Val Lys Gly Glu Arg Val Ser Phe 315 320 325	1073
CCT GTA TGC ACC TAC GTC CCC TCA ACC ATC TGT GAT CAA ATG ACT GGC Pro Val Cys Thr Tyr Val Pro Ser Thr Ile Cys Asp Gln Met Thr Gly 330 335 340 345	1121
ATA CTA GCG ACC GAC GTC ACA CCG GAG GAC GCA CAG AAG TTG TTA GTG Ile Leu Ala Thr Asp Val Thr Pro Glu Asp Ala Gln Lys Leu Leu Val 350 355 360	1169
GCA TTG AAT CAG AGG ATA GTT GTG AAC GGA AGA ACA CAG CGA AAC ACT Gly Leu Asn Gln Arg Ile Val Val Asn Gly Arg Thr Gln Arg Asn Thr 365 370 375	1217
AAC ACG ATG AAG AAC TAT CTG CTT CCG ATT GTG GCC GTC GCA TTT AGC Asn Thr Met Lys Asn Tyr Leu Leu Pro Ile Val Ala Val Ala Phe Ser 380 385 390	1265
AAG TGG GCG AGG GAA TAC AAG GCA GAC CTT GAT GAT GAA AAA CCT CTG Lys Trp Ala Arg Glu Tyr Lys Ala Asp Leu Asp Asp Glu Lys Pro Leu 395 400 405	1313
GCT GTC CGA GAG AGG TCA CTT ACT TGC TGC TGC TTG TGG GCA TTT AAA Gly Val Arg Glu Arg Ser Leu Thr Cys Cys Cys Leu Trp Ala Phe Lys 410 415 420 425	1361
ACG AGG AAG ATG CAC ACC ATG TAC AAG AAA CCA GAC ACC CAG ACA ATA Thr Arg Lys Met His Thr Met Tyr Lys Lys Pro Asp Thr Gln Thr Ile 430 435 440	1409
GTG AAG GTG CCT TCA GAG TTT AAC TCG TTC GTC ATC CCG AGC CTA TGG Val Lys Val Pro Ser Glu Phe Asn Ser Phe Val Ile Pro Ser Leu Trp 445 450 455	1457
TCT ACA GGC CTC GCA ATC CCA GTC AGA TCA CGC ATT AAG ATG CTT TTG Ser Thr Gly Leu Ala Ile Pro Val Arg Ser Arg Ile Lys Met Leu Leu 460 465 470	1505
GCC AAG AAG ACC AAG CGA GAG TTA ATA CCT GTT CTC GAC GCG TCG TCA Ala Lys Lys Thr Lys Arg Glu Leu Ile Pro Val Leu Asp Ala Ser Ser 475 480 485	1553

GCC AGG GAT GCT GAA CAA GAG GAG AAG GAG AGG TTG GAG GCC GAG CTG Ala Arg Asp Ala Glu Gln Glu Glu Lys Glu Arg Leu Glu Ala Glu Leu 490 495 500 505	1601
ACT AGA GAA GCC TTA CCA CCC CTC GTC CCC ATC GCG CCG GCG GAG ACG Thr Arg Glu Ala Leu Pro Pro Leu Val Pro Ile Ala Pro Ala Glu Thr 510 515 520	1649
GGA GTC GTC GAC GTC GAC GTT GAA GAA CTA GAG TAT CAC GCA GGT GCA Gly Val Val Asp Val Asp Val Glu Glu Leu Glu Tyr His Ala Gly Ala 525 530 535	1697
<i>G</i> GGG GTC GTG GAA ACA CCT CGC AGC GCG TTG AAA GTC ACC GCA CAG CCG Gly Val Val Glu Thr Pro Arg Ser Ala Leu Lys Val Thr Ala Gln Pro 540 545 550	1745
AAC GAC GTA CTA CTA GGA AAT TAC GTA GTT CTG TCC CCG CAG ACC GTG Asn Asp Val Leu Leu Gly Asn Tyr Val Val Leu Ser Pro Gln Thr Val 555 560 565	1793
<i>C</i> C AAG AGC TCC AAG TTG GCC CCC GTG CAC CCT CTA GCA GAG CAG GTG Leu Lys Ser Ser Lys Leu Ala Pro Val His Pro Leu Ala Glu Gln Val 570 575 580 585	1841
<i>A</i> ATA ATA ACA CAT AAC GGG AGG GCC GGC GGT TAC CAG GTC GAC GGA Lys Ile Ile Thr His Asn Gly Arg Ala Gly Gly Tyr Gln Val Asp Gly 590 595 600	1889
<i>T</i> TAT GAC GGC AGG GTC CTA CTA CCA TGT GGA TCG GCC ATT CCG GTC CCT Tyr Asp Gly Arg Val Leu Leu Pro Cys Gly Ser Ala Ile Pro Val Pro 605 610 615	1937
<i>G</i> GAG TTT CAA GCT TTG AGC GAG AGC GCC ACT ATG GTG TAC AAC GAA AGG Glu Phe Gln Ala Leu Ser Glu Ser Ala Thr Met Val Tyr Asn Glu Arg 620 625 630	1985
GAG TTC GTC AAC AGG AAA CTA TAC CAT ATT GCC GTT CAC GGA CCG TCG Glu Phe Val Asn Arg Lys Leu Tyr His Ile Ala Val His Gly Pro Ser 635 640 645	2033
CTG AAC ACC GAC GAG GAG AAC TAC GAG AAA GTC AGA GCT GAA AGA ACT Leu Asn Thr Asp Glu Glu Asn Tyr Glu Lys Val Arg Ala Glu Arg Thr 650 655 660 665	2081
GAC GCC GAG TAC GTG TTC GAC GTA GAT AAA AAA TGC TGC GTC AAG AGA Asp Ala Glu Tyr Val Phe Asp Val Asp Lys Lys Cys Cys Val Lys Arg 670 675 680	2129
GAG GAA GCG TCG GGT TTG GTG TTG GTG GGA GAG CTA ACC AAC CCC CCG Glu Glu Ala Ser Gly Leu Val Leu Val Gly Glu Leu Thr Asn Pro Pro 685 690 695	2177

TTC CAT GAA TTC GCC TAC GAA GGG CTG AAG ATC AGG CCG TCG GCA CCA Phe His Glu Phe Ala Tyr Glu Gly Leu Lys Ile Arg Pro Ser Ala Pro 700 705 710	2225
TAT AAG ACT ACA GTA GTA GGA GTC TTT GGG GTT CCG GGA TCA GGC AAG Tyr Lys Thr Thr Val Val Gly Val Phe Gly Val Pro Gly Ser Gly Lys 715 720 725	2273
TCT GCT ATT ATT AAG AGC CTC GTG ACC AAA CAC GAT CTG GTC ACC AGC Ser Ala Ile Ile Lys Ser Leu Val Thr Lys His Asp Leu Val Thr Ser 730 735 740 745	2321
GGC AAG AAG GAG AAC TGC CAG GAA ATA GTT AAC GAC GTG AAG AAG CAC Gly Lys Lys Glu Asn Cys Gln Glu Ile Val Asn Asp Val Lys Lys His 750 755 760	2369
CGC GGG AAG GGG ACA AGT AGG GAA AAC AGT GAC TCC ATC CTG CTA AAC Arg Gly Lys Gly Thr Ser Arg Glu Asn Ser Asp Ser Ile Leu Leu Asn 765 770 775	2417
GGG TGT CGT CGT GCC GTG GAC ATC CTA TAT GTG GAC GAG GCT TTC GCT Gly Cys Arg Arg Ala Val Asp Ile Leu Tyr Val Asp Glu Ala Phe Ala 780 785 790	2465
GCG CAT TCC GGT ACT CTG CTG GCC CTA ATT GCT CTT GTT AAA CCT CGG Gly His Ser Gly Thr Leu Leu Ala Leu Ile Ala Leu Val Lys Pro Arg 795 800 805	2513
AGC AAA GTG GTG TTA TGC GGA GAC CCC AAG CAA TGC GGA TTC TTC AAT Ser Lys Val Val Leu Cys Gly Asp Pro Lys Gln Cys Gly Phe Phe Asn 810 815 820 825	2561
ATG ATG CAG CTT AAG GTG AAC TTC AAC CAC AAC ATC TGC ACT GAA GTA Met Met Gln Leu Lys Val Asn Phe Asn His Asn Ile Cys Thr Glu Val 830 835 840	2609
TGT CAT AAA AGT ATA TCC AGA CGT TGC ACG CGT CCA GTC ACG GCC ATC Cys His Lys Ser Ile Ser Arg Arg Cys Thr Arg Pro Val Thr Ala Ile 845 850 855	2657
GTG TCT ACG TTG CAC TAC GGA GGC AAG ATG CGC ACG ACC AAC CCG TGC Val Ser Thr Leu His Tyr Gly Gly Lys Met Arg Thr Thr Asn Pro Cys 860 865 870	2705
AAC AAA CCC ATA ATC ATA GAC ACC ACA GGA CAG ACC AAG CCC AAG CCA Asn Lys Pro Ile Ile Ile Asp Thr Thr Gly Gln Thr Lys Pro Lys Pro 875 880 885	2753
GGA GAC ATC GTG TTA ACA TGC TTC CGA GGC TGG GCA AAG CAG CTG CAG Gly Asp Ile Val Leu Thr Cys Phe Arg Gly Trp Ala Lys Gln Leu Gln 890 895 900 905	2801

TTG GAC TAC CGT GGA CAC GAA GTC ATG ACA GCA GCA GCA TCT CAG GGC Leu Asp Tyr Arg Gly His Glu Val Met Thr Ala Ala Ala Ser Gln Gly 910	915	920	2849
CTC ACC CGC AAA GGG GTA TAC GCC GTA AGG CAG AAG GTG AAT GAA AAT Leu Thr Arg Lys Gly Val Tyr Ala Val Arg Gln Lys Val Asn Glu Asn 925	930	935	2897
CCC TTG TAT GCC CCT GCG TCG GAG CAC GTG AAT GTA CTG CTG ACG CGC Pro Leu Tyr Ala Pro Ala Ser Glu His Val Asn Val Leu Leu Thr Arg 940	945	950	2945
ACT GAG GAT AGG CTG GTG TGG AAA ACG CTG GCC GGC GAT CCC TGG ATT Thr Glu Asp Arg Leu Val Trp Lys Thr Leu Ala Gly Asp Pro Trp Ile 955	960	965	2993
AAG GTC CTA TCA AAC ATT CCA CAG GGT AAC TTT ACG GCC ACA TTG GAA Lys Val Leu Ser Asn Ile Pro Gln Gly Asn Phe Thr Ala Thr Leu Glu 970	975	980	3041
GAA TGG CAA GAA GAA CAC GAC AAA ATA ATG AAG GTG ATT GAA GGA CCG Glu Trp Gln Glu Glu His Asp Lys Ile Met Lys Val Ile Glu Gly Pro 990	995	1000	3089
GCT GCG CCT GTG GAC GCG TTC CAG AAC AAA GCG AAC GTG TGT TGG GCG Ala Ala Pro Val Asp Ala Phe Gln Asn Lys Ala Asn Val Cys Trp Ala 1005	1010	1015	3137
AAA AGC CTG GTG CCT GTC CTG GAC ACT GCC GGA ATC AGA TTG ACA GCA Lys Ser Leu Val Pro Val Leu Asp Thr Ala Gly Ile Arg Leu Thr Ala 1020	1025	1030	3185
GAG GAG TGG AGC ACC ATA ATT ACA GCA TTT AAG GAG GAC AGA GCT TAC Glu Glu Trp Ser Thr Ile Ile Thr Ala Phe Lys Glu Asp Arg Ala Tyr 1035	1040	1045	3233
TCT CCA GTG GTG GCC TTG AAT GAA ATT TGC ACC AAG TAC TAT GGA GTT Ser Pro Val Val Ala Leu Asn Glu Ile Cys Thr Lys Tyr Tyr Gly Val 1050	1055	1060	3281
GAC CTG GAC AGT GGC CTG TTT TCT GCC CCG AAG GTG TCC CTG TAT TAC Asp Leu Asp Ser Gly Leu Phe Ser Ala Pro Lys Val Ser Leu Tyr Tyr 1070	1075	1080	3329
GAG AAC AAC CAC TGG GAT AAC AGA CCT GGT GGA AGG ATG TAT GGA TTC Glu Asn Asn His Trp Asp Asn Arg Pro Gly Gly Arg Met Tyr Gly Phe 1085	1090	1095	3377
AAT GCC GCA ACA GCT GCC AGG CTG GAA GCT AGA CAT ACC TTC CTG AAG Asn Ala Ala Thr Ala Ala Arg Leu Glu Ala Arg His Thr Phe Leu Lys 1100	1105	1110	3425

GGG CAG TGG CAT ACG GGC AAG CAG GCA GTT ATC GCA GAA AGA AAA ATC Gly Gln Trp His Thr Gly Lys Gln Ala Val Ile Ala Glu Arg Lys Ile 1115 1120 1125	3473
CAA CCG CTT TCT GTG CTG GAC AAT GTA ATT CCT ATC AAC CGC AGG CTG Gln Pro Leu Ser Val Leu Asp Asn Val Ile Pro Ile Asn Arg Arg Leu 1130 1135 1140 1145	3521
CCG CAC GCC CTG GTG GCT GAG TAC AAG ACG GTT AAA GGC AGT AGG GTT Pro His Ala Leu Val Ala Glu Tyr Lys Thr Val Lys Gly Ser Arg Val 1150 1155 1160	3569
GAG TGG CTG GTC AAT AAA GTA AGA GGG TAC CAC GTC CTG CTG GTG AGT Glu Trp Leu Val Asn Lys Val Arg Gly Tyr His Val Leu Leu Val Ser 1165 1170 1175	3617
GAG TAC AAC CTG GCT TTG CCT CGA CGC AGG GTC ACT TGG TTG TCA CCG Glu Tyr Asn Leu Ala Leu Pro Arg Arg Arg Val Thr Trp Leu Ser Pro 1180 1185 1190	3665
CTG AAT GTC ACA GGC GCC GAT AGG TGC TAC GAC CTA AGT TTA GGA CTG Leu Asn Val Thr Gly Ala Asp Arg Cys Tyr Asp Leu Ser Leu Gly Leu 1195 1200 1205	3713
GCG GCT GAC GCC GGC AGG TTC GAC TTG GTC TTT GTG AAC ATT CAC ACG Pro Ala Asp Ala Gly Arg Phe Asp Leu Val Phe Val Asn Ile His Thr 1210 1215 1220 1225	3761
GAA TTC AGA ATC CAC CAC TAC CAG CAG TGT GTC GAC CAC GCC ATG AAG Glu Phe Arg Ile His His Tyr Gln Gln Cys Val Asp His Ala Met Lys 1230 1235 1240	3809
CTG CAG ATG CTT GGG GGA GAT GCG CTA CGA CTG CTA AAA CCC GGC GGC Leu Gln Met Leu Gly Gly Asp Ala Leu Arg Leu Leu Lys Pro Gly Gly 1245 1250 1255	3857
ATC TTG ATG AGA GCT TAC GGA TAC GCC GAT AAA ATC AGC GAA GCC GTT Ile Leu Met Arg Ala Tyr Gly Tyr Ala Asp Lys Ile Ser Glu Ala Val 1260 1265 1270	3905
GTT TCC TCC TTA AGC AGA AAG TTC TCG TCT GCA AGA GTG TTG CGC CCG Val Ser Ser Leu Ser Arg Lys Phe Ser Ser Ala Arg Val Leu Arg Pro 1275 1280 1285	3953
GAT TGT GTC ACC AGC AAT ACA GAA GTG TTC TTG CTG TTC TCC AAC TTT Asp Cys Val Thr Ser Asn Thr Glu Val Phe Leu Leu Phe Ser Asn Phe 1290 1295 1300 1305	4001
GAC AAC GGA AAG AGA CCC TCT ACG CTA CAC CAG ATG AAT ACC AAG CTG Asp Asn Gly Lys Arg Pro Ser Thr Leu His Gln Met Asn Thr Lys Leu 1310 1315 1320	4049

AGT	GCC	GTG	TAT	GCC	GGA	GAA	GCC	ATG	CAC	ACG	GCC	GGG	TGT	GCA	CCA		4097
Ser	Ala	Val	Tyr	Ala	Gly	Glu	Ala	Met	His	Thr	Ala	Gly	Cys	Ala	Pro		
1325								1330						1335			
TCC	TAC	AGA	GTT	AAG	AGA	GCA	GAC	ATA	GCC	ACG	TGC	ACA	GAA	GCG	GCT		4145
Ser	Tyr	Arg	Val	Lys	Arg	Ala	Asp	Ile	Ala	Thr	Cys	Thr	Glu	Ala	Ala		
1340								1345					1350				
GTG	GTT	AAC	GCA	GCT	AAC	GCC	CGT	GGA	ACT	GTA	GGG	GAT	GGC	GTA	TGC		4193
Val	Val	Asn	Ala	Ala	Asn	Ala	Arg	Gly	Thr	Val	Gly	Asp	Gly	Val	Cys		
1355								1360					1365				
AGG	GCC	GTG	GCG	AAG	AAA	TGG	CCG	TCA	GCC	TTT	AAG	GGA	GCA	GCA	ACA		4241
Arg	Ala	Val	Ala	Lys	Lys	Trp	Pro	Ser	Ala	Phe	Lys	Gly	Ala	Ala	Thr		
1370								1375					1380			1385	
CCA	GTG	GGC	ACA	ATT	AAA	ACA	GTC	ATG	TGC	GGC	TCG	TAC	CCC	GTC	ATC		4289
Pro	Val	Gly	Thr	Ile	Lys	Thr	Val	Met	Cys	Gly	Ser	Tyr	Pro	Val	Ile		
1390								1395					1400				
CAC	GCT	GTA	GCG	CCT	AAT	TTC	TCT	GCC	ACG	ACT	GAA	GCG	GAA	GGG	GAC		4337
His	Ala	Val	Ala	Pro	Asn	Phe	Ser	Ala	Thr	Thr	Glu	Ala	Glu	Gly	Asp		
1405								1410					1415				
CGC	GAA	TTG	GCC	GCT	GTC	TAC	CGG	GCA	GTG	GCC	GCC	GAA	GTA	AAC	AGA		4385
Arg	Glu	Leu	Ala	Ala	Val	Tyr	Arg	Ala	Val	Ala	Ala	Glu	Val	Asn	Arg		
1420								1425					1430				
GTG	TCA	CTG	AGC	AGC	GTA	GCC	ATC	CCG	CTG	CTG	TCC	ACA	GGA	GTG	TTC		4433
Leu	Ser	Leu	Ser	Ser	Val	Ala	Ile	Pro	Leu	Leu	Ser	Thr	Gly	Val	Phe		
1435								1440					1445				
AGC	GGC	GGA	AGA	GAT	AGG	CTG	CAG	CAA	TCC	CTC	AAC	CAT	CTA	TTC	ACA		4481
Ser	Gly	Gly	Arg	Asp	Arg	Leu	Gln	Gln	Ser	Leu	Asn	His	Leu	Phe	Thr		
1450								1455					1460			1465	
GCA	ATG	GAC	GCC	ACG	GAC	GCT	GAC	GTG	ACC	ATC	TAC	TGC	AGA	GAC	AAA		4529
Ala	Met	Asp	Ala	Thr	Asp	Ala	Asp	Val	Thr	Ile	Tyr	Cys	Arg	Asp	Lys		
1470								1475					1480				
AGT	TGG	GAG	AAG	AAA	ATC	CAG	GAA	GCC	ATT	GAC	ATG	AGG	ACG	GCT	GTG		4577
Ser	Trp	Glu	Lys	Lys	Ile	Gln	Glu	Ala	Ile	Asp	Met	Arg	Thr	Ala	Val		
1485								1490					1495				
GAG	TTG	CTC	AAT	GAT	GAC	GTG	GAG	CTG	ACC	ACA	GAC	TTG	GTG	AGA	GTG		4625
Glu	Leu	Leu	Asn	Asp	Asp	Val	Glu	Leu	Thr	Thr	Asp	Leu	Val	Arg	Val		
1500								1505					1510				
CAC	CCG	GAC	AGC	AGC	CTG	GTG	GGT	CGT	AAG	GGC	TAC	AGT	ACC	ACT	GAC		4673
His	Pro	Asp	Ser	Ser	Leu	Val	Gly	Arg	Lys	Gly	Tyr	Ser	Thr	Thr	Asp		
1515								1520					1525				

GGG TCG CTG TAC TCG TAC TTT GAA GGT ACG AAA TTC AAC CAG GCT GCT Gly Ser Leu Tyr Ser Tyr Phe Glu Gly Thr Lys Phe Asn Gln Ala Ala 1530 1535 1540 1545	4721
ATT GAT ATG GCA GAG ATA CTG ACG TTG TGG CCC AGA CTG CAA GAG GCA Ile Asp Met Ala Glu Ile Leu Thr Leu Trp Pro Arg Leu Gln Glu Ala 1550 1555 1560	4769
AAC GAA CAG ATA TGC CTA TAC GCG CTG GGC GAA ACA ATG GAC AAC ATC Asn Glu Gln Ile Cys Leu Tyr Ala Leu Gly Glu Thr Met Asp Asn Ile 1565 1570 1575	4817
AGA TCC AAA TGT CCG GTG AAC GAT TCC GAT TCA TCA ACA CCT CCC AGG Arg Ser Lys Cys Pro Val Asn Asp Ser Asp Ser Ser Thr Pro Pro Arg 1580 1585 1590	4865
ACA GTG CCC TGC CTG TGC CGC TAC GCA ATG ACA GCA GAA CGG ATC GCC Thr Val Pro Cys Leu Cys Arg Tyr Ala Met Thr Ala Glu Arg Ile Ala 1595 1600 1605	4913
GCG CTT AGG TCA CAC CAA GTT AAA AGC ATG GTG GTT TGC TCA TCT TTT Arg Leu Arg Ser His Gln Val Lys Ser Met Val Val Cys Ser Ser Phe 1610 1615 1620 1625	4961
GCC CTC CCG AAA TAC CAT GTA GAT GGG GTG CAG AAG GTA AAG TGC GAG Pro Leu Pro Lys Tyr His Val Asp Gly Val Gln Lys Val Lys Cys Glu 1630 1635 1640	5009
AGG GTT CTC CTG TTC GAC CCG ACG GTA CCT TCA GTG GTT AGT CCG CGG Lys Val Leu Leu Phe Asp Pro Thr Val Pro Ser Val Val Ser Pro Arg 1645 1650 1655	5057
AAG TAT GCC GCA TCT ACG ACG GAC CAC TCA GAT CGG TCG TTA CGA GGG Lys Tyr Ala Ala Ser Thr Thr Asp His Ser Asp Arg Ser Leu Arg Gly 1660 1665 1670	5105
TTT GAC TTG GAC TGG ACC ACC GAC TCG TCT TCC ACT GCC AGC GAT ACC Phe Asp Leu Asp Trp Thr Thr Asp Ser Ser Ser Thr Ala Ser Asp Thr 1675 1680 1685	5153
ATG TCG CTA CCC AGT TTG CAG TCG TGT GAC ATC GAC TCG ATC TAC GAG Met Ser Leu Pro Ser Leu Gln Ser Cys Asp Ile Asp Ser Ile Tyr Glu 1690 1695 1700 1705	5201
CCA ATG GCT CCC ATA GTA GTG ACG GCT GAC GTA CAC CCT GAA CCC GCA Pro Met Ala Pro Ile Val Val Thr Ala Asp Val His Pro Glu Pro Ala 1710 1715 1720	5249
GGC ATC GCG GAC CTG GCG GCA GAT GTG CAC CCT GAA CCC GCA GAC CAT Gly Ile Ala Asp Leu Ala Ala Asp Val His Pro Glu Pro Ala Asp His 1725 1730 1735	5297

GTG GAC CTC GAG AAC CCG ATT CCT CCA CCG CGC CCG AAG AGA GCT GCA Val Asp Leu Glu Asn Pro Ile Pro Pro Pro Arg Pro Lys Arg Ala Ala 1740 1745 1750	5345
TAC CTT GCC TCC CGC GCG GCG GAG CGA CCG GTG CCG GCG CCG AGA AAG Tyr Leu Ala Ser Arg Ala Ala Glu Arg Pro Val Pro Ala Pro Arg Lys 1755 1760 1765	5393
CCG ACG CCT GCC CCA AGG ACT GCG TTT AGG AAC AAG CTG CCT TTG ACG Pro Thr Pro Ala Pro Arg Thr Ala Phe Arg Asn Lys Leu Pro Leu Thr 1770 1775 1780 1785	5441
TTC GGC GAC TTT GAC GAG CAC GAG GTC GAT GCG TTG GCC TCC GGG ATT Phe Gly Asp Phe Asp Glu His Glu Val Asp Ala Leu Ala Ser Gly Ile 1790 1795 1800	5489
ACT TTC GGA GAC TTC GAC GAC GTC CTG CGA CTA GGC CGC GCG GGT GCA Thr Phe Gly Asp Phe Asp Asp Val Leu Arg Leu Gly Arg Ala Gly Ala 1805 1810 1815	5537
TAT ATT TTC TCC TCG GAC ACT GGC AGC GGA CAT TTA CAA CAA AAA TCC Tyr Ile Phe Ser Ser Asp Thr Gly Ser Gly His Leu Gln Gln Lys Ser 1820 1825 1830	5585
GTT AGG CAG CAC AAT CTC CAG TGC GCA CAA CTG GAT GCG GTC CAG GAG Val Arg Gln His Asn Leu Gln Cys Ala Gln Leu Asp Ala Val Gln Glu 1835 1840 1845	5633
GAG AAA ATG TAC CCG CCA AAA TTG GAT ACT GAG AGG GAG AAG CTG TTG Glu Lys Met Tyr Pro Pro Lys Leu Asp Thr Glu Arg Glu Lys Leu Leu 1850 1855 1860 1865	5681
CTG CTG AAA ATG CAG ATG CAC CCA TCG GAG GCT AAT AAG AGT CGA TAC Leu Leu Lys Met Gln Met His Pro Ser Glu Ala Asn Lys Ser Arg Tyr 1870 1875 1880	5729
CAG TCT CGC AAA GTG GAG AAC ATG AAA GCC ACG GTG GTG GAC AGG CTC Gln Ser Arg Lys Val Glu Asn Met Lys Ala Thr Val Val Asp Arg Leu 1885 1890 1895	5777
ACA TCG GGG GCC AGA TTG TAC ACG GGA GCG GAC GTA GGC CGC ATA CCA Thr Ser Gly Ala Arg Leu Tyr Thr Gly Ala Asp Val Gly Arg Ile Pro 1900 1905 1910	5825
ACA TAC GCG GTT CGG TAC CCC CGC CCC GTG TAC TCC CCT ACC GTG ATC Thr Tyr Ala Val Arg Tyr Pro Arg Pro Val Tyr Ser Pro Thr Val Ile 1915 1920 1925	5873
GAA AGA TTC TCA AGC CCC GAT GTA GCA ATC GCA GCG TGC AAC GAA TAC Glu Arg Phe Ser Ser Pro Asp Val Ala Ile Ala Ala Cys Asn Glu Tyr 1930 1935 1940 1945	5921

CTA TCC AGA AAT TAC CCA ACA GTG GCG TCG TAC CAG ATA ACA GAT GAA Leu Ser Arg Asn Tyr Pro Thr Val Ala Ser Tyr Gln Ile Thr Asp Glu 1950 1955 1960	5969
TAC GAC GCA TAC TTG GAC ATG GTT GAC GGG TCG GAT AGT TGC TTG GAC Tyr Asp Ala Tyr Leu Asp Met Val Asp Gly Ser Asp Ser Cys Leu Asp 1965 1970 1975	6017
AGA GCG ACA TTC TGC CCG GCG AAG CTC CGG TGC TAC CCG AAA CAT CAT Arg Ala Thr Phe Cys Pro Ala Lys Leu Arg Cys Tyr Pro Lys His His 1980 1985 1990	6065
GCG TAC CAC CAG CCG ACT GTA CGC AGT GCC GTC CCG TCA CCC TTT CAG Ala Tyr His Gln Pro Thr Val Arg Ser Ala Val Pro Ser Pro Phe Gln 1995 2000 2005	6113
AAC ACA CTA CAG AAC GTG CTA GCG GCC ACC AAG AGA AAC TGC AAC Asn Thr Leu Gln Asn Val Leu Ala Ala Thr Lys Arg Asn Cys Asn 2010 2015 2020 2025	6161
GTC ACG CAA ATG CGA GAA CTA CCC ACC ATG GAC TCG GCA GTG TTC AAC Val Thr Gln Met Arg Glu Leu Pro Thr Met Asp Ser Ala Val Phe Asn 2030 2035 2040	6209
GTC GAG TGC TTC AAG CGC TAT GCC TGC TCC GGA GAA TAT TGG GAA GAA Val Glu Cys Phe Lys Arg Tyr Ala Cys Ser Gly Glu Tyr Trp Glu Glu 2045 2050 2055	6257
TAT GCT AAA CAA CCT ATC CGG ATA ACC ACT GAG AAC ATC ACT ACC TAT Tyr Ala Lys Gln Pro Ile Arg Ile Thr Thr Glu Asn Ile Thr Thr Tyr 2060 2065 2070	6305
GTC ACC AAA TTG AAA GGC CCG AAA GCT GCT GCC TTG TTC GCT AAG ACC Val Thr Lys Leu Lys Gly Pro Lys Ala Ala Leu Phe Ala Lys Thr 2075 2080 2085	6353
CAC AAC TTG GTT CCG CTG CAG GAG GTT CCC ATG GAC AGA TTC ACG GTC His Asn Leu Val Pro Leu Gln Glu Val Pro Met Asp Arg Phe Thr Val 2090 2095 2100 2105	6401
GAC ATG AAA CGA GAT GTC AAA GTC ACT CCA GGG ACG AAA CAC ACA GAG Asp Met Lys Arg Asp Val Lys Val Thr Pro Gly Thr Lys His Thr Glu 2110 2115 2120	6449
GAA AGA CCC AAA GTC CAG GTA ATT CAA GCA GCG GAG CCA TTG GCG ACC Glu Arg Pro Lys Val Gln Val Ile Gln Ala Ala Glu Pro Leu Ala Thr 2125 2130 2135	6497
GCT TAC CTG TGC GGC ATC CAC AGG GAA TTA GTA AGG AGA CTA AAT GCT Ala Tyr Leu Cys Gly Ile His Arg Glu Leu Val Arg Arg Leu Asn Ala 2140 2145 2150	6545

GTG TTA CGC CCT AAC GTG CAC ACA TTG TTT GAT ATG TCG GCC GAA GAC Val Leu Arg Pro Asn Val His Thr Leu Phe Asp Met Ser Ala Glu Asp 2155 2160 2165	6593
TTT GAC GCG ATC ATC GCC TCT CAC TTC CAC CCA GGA GAC CCG GTT CTA Phe Asp Ala Ile Ile Ala Ser His Phe His Pro Gly Asp Pro Val Leu 2170 2175 2180 2185	6641
GAG ACG GAC ATT GCA TCA TTC GAC AAA AGC CAG GAC GAC TCC TTG GCT Glu Thr Asp Ile Ala Ser Phe Asp Lys Ser Gln Asp Asp Ser Leu Ala 2190 2195 2200	6689
CTT ACA GGT TTA ATG ATC CTC GAA GAT CTA GGG GTG GAT CAG TAC CTG Leu Thr Gly Leu Met Ile Leu Glu Asp Leu Gly Val Asp Gln Tyr Leu 2205 2210 2215	6737
CTG GAC TTG ATC GAG GCA GCC TTT GGG GAA ATA TCC AGC TGT CAC CTA Leu Asp Leu Ile Glu Ala Ala Phe Gly Glu Ile Ser Ser Cys His Leu 2220 2225 2230	6785
CCA ACT GGC ACG CGC TTC AAG TTC GGA GCT ATG ATG AAA TCG GGC ATG Pro Thr Gly Thr Arg Phe Lys Phe Gly Ala Met Met Lys Ser Gly Met 2235 2240 2245	6833
TTT CTG ACT TTG TTT ATT AAC ACT GTT TTG AAC ATC ACC ATA GCA AGC Phe Leu Thr Leu Phe Ile Asn Thr Val Leu Asn Ile Thr Ile Ala Ser 2250 2255 2260 2265	6881
AGG GTA CTG GAG CAG AGA CTC ACT GAC TCC GCC TGT GCG GCC TTC ATC Arg Val Leu Glu Gln Arg Leu Thr Asp Ser Ala Cys Ala Ala Phe Ile 2270 2275 2280	6929
GGC GAC GAC AAC ATC GTT CAC GGA GTG ATC TCC GAC AAG CTG ATG GCG Gly Asp Asp Asn Ile Val His Gly Val Ile Ser Asp Lys Leu Met Ala 2285 2290 2295	6977
GAG AGG TGC GCG TCG TGG GTC AAC ATG GAG GTG AAG ATC ATT GAC GCT Glu Arg Cys Ala Ser Trp Val Asn Met Glu Val Lys Ile Ile Asp Ala 2300 2305 2310	7025
GTC ATG GGC GAA AAA CCC CCA TAT TTT TGT GGG GGA TTC ATA GTT TTT Val Met Gly Glu Lys Pro Pro Tyr Phe Cys Gly Gly Phe Ile Val Phe 2315 2320 2325	7073
GAC AGC GTC ACA CAG ACC GCC TGC CGT GTT TCA GAC CCA CTT AAG CGC Asp Ser Val Thr Gln Thr Ala Cys Arg Val Ser Asp Pro Leu Lys Arg 2330 2335 2340 2345	7121
CTG TTC AAG TTG GGT AAG CCG CTA ACA GCT GAA GAC AAG CAG GAC GAA Leu Phe Lys Leu Gly Lys Pro Leu Thr Ala Glu Asp Lys Gln Asp Glu 2350 2355 2360	7169

GAC AGG CGA CGA GCA CTG AGT GAC GAG GTT AGC AAG TGG TTC CGG ACA Asp Arg Arg Arg Ala Leu Ser Asp Glu Val Ser Lys Trp Phe Arg Thr 2365 2370 2375	7217
GGC TTG GGG GCC GAA CTG GAG GTG GCA CTA ACA TCT AGG TAT GAG GTA Gly Leu Gly Ala Glu Leu Glu Val Ala Leu Thr Ser Arg Tyr Glu Val 2380 2385 2390	7265
GAG GGC TGC AAA AGT ATC CTC ATA GCC ATG ACC ACC TTG GCG AGG GAC Glu Gly Cys Lys Ser Ile Leu Ile Ala Met Thr Thr Leu Ala Arg Asp 2395 2400 2405	7313
ATT AAG GCG TTT AAG AAA TTG AGA GGA CCT GTT ATA CAC CTC TAC GGC Ile Lys Ala Phe Lys Lys Leu Arg Gly Pro Val Ile His Leu Tyr Gly 2410 2415 2420 2425	7361
GGT CCT AGA TTG GTG CGT TAATACACAG AATTCTGATT ATAGCGCACT Gly Pro Arg Leu Val Arg 2430	7409
ATTATAGCAC C ATG AAT TAC ATC CCT ACG CAA ACG TTT TAC GGC CGC CGG Met Asn Tyr Ile Pro Thr Gln Thr Phe Tyr Gly Arg Arg 1 5 10	7459
GG CGC CCG CGC CCG GCG GCC CGT CCT TGG CCG TTG CAG GCC ACT CCG Trp Arg Pro Arg Pro Ala Ala Arg Pro Trp Pro Leu Gln Ala Thr Pro 15 20 25	7507
G TG GCT CCC GTC GTC CCC GAC TTC CAG GCC CAG CAG ATG CAG CAA CTC Val Ala Pro Val Val Pro Asp Phe Gln Ala Gln Gln Met Gln Gln Leu 30 35 40 45	7555
ATC AGC GCC GTA AAT GCG CTG ACA ATG AGA CAG AAC GCA ATT GCT CCT Ile Ser Ala Val Asn Ala Leu Thr Met Arg Gln Asn Ala Ile Ala Pro 50 55 60	7603
GCT AGG CCT CCC AAA CCA AAG AAG AAG AAG ACA ACC AAA CCA AAG CCG Ala Arg Pro Pro Lys Pro Lys Lys Lys Thr Thr Lys Pro Lys Pro 65 70 75	7651
AAA ACG CAG CCC AAG AAG ATC AAC GGA AAA ACG CAG CAG CAA AAG AAG Lys Thr Gln Pro Lys Lys Ile Asn Gly Lys Thr Gln Gln Lys Lys 80 85 90	7699
AAA GAC AAG CAA GCC GAC AAG AAG AAG AAG AAA CCC GGA AAA AGA GAA Lys Asp Lys Gln Ala Asp Lys Lys Lys Lys Pro Gly Lys Arg Glu 95 100 105	7747
AGA ATG TGC ATG AAG ATT GAA AAT GAC TGT ATC TTC GAA GTC AAA CAC Arg Met Cys Met Lys Ile Glu Asn Asp Cys Ile Phe Glu Val Lys His 110 115 120 125	7795

GAA GGA AAG GTC ACT GGG TAC GCC TGC CTG GTG GGC GAC AAA GTC ATG Glu Gly Lys Val Thr Gly Tyr Ala Cys Leu Val Gly Asp Lys Val Met 130 135 140	7843
AAA CCT GCC CAC GTG AAA GGA GTC ATC GAC AAC GCG GAC CTG GCA AAG Lys Pro Ala His Val Lys Gly Val Ile Asp Asn Ala Asp Leu Ala Lys 145 150 155	7891
CTA GCT TTC AAG AAA TCG AGC AAG TAT GAC CTT GAG TGT GCC CAG ATA Leu Ala Phe Lys Lys Ser Ser Lys Tyr Asp Leu Glu Cys Ala Gln Ile 160 165 170	7939
CCA GTT CAC ATG AGG TCG GAT GCC TCA AAG TAC ACG CAT GAG AAG CCC Pro Val His Met Arg Ser Asp Ala Ser Lys Tyr Thr His Glu Lys Pro 175 180 185	7987
GAG GGA CAC TAT AAC TGG CAC CAC GGG GCT GTT CAG TAC AGC GGA GGT Glu Gly His Tyr Asn Trp His His Gly Ala Val Gln Tyr Ser Gly Gly 190 195 200 205	8035
AGG TTC ACT ATA CCG ACA GGA GCG GGC AAA CCG GGA GAC AGT GGC CGG Arg Phe Thr Ile Pro Thr Gly Ala Gly Lys Pro Gly Asp Ser Gly Arg 210 215 220	8083
CCC ATC TTT GAC AAC AAG GGG AGG GTA GTC GCT ATC GTC CTG GGC GGG Pro Ile Phe Asp Asn Lys Gly Arg Val Val Ala Ile Val Leu Gly Gly 225 230 235	8131
GCC AAC GAG GGC TCA CGC ACA GCA CTG TCG GTG GTC ACC TGG AAC AAA Ala Asn Glu Gly Ser Arg Thr Ala Leu Ser Val Val Thr Trp Asn Lys 240 245 250	8179
GAT ATG GTG ACT AGA GTG ACC CCC GAG GGG TCC GAA GAG TGG TCC GCC Asp Met Val Thr Arg Val Thr Pro Glu Gly Ser Glu Glu Trp Ser Ala 255 260 265	8227
CCG CTG ATT ACT GCC ATG TGT GTC CTT GCC AAT GCT ACC TTC CCG TGC Pro Leu Ile Thr Ala Met Cys Val Leu Ala Asn Ala Thr Phe Pro Cys 270 275 280 285	8275
TTC CAG CCC CCG TGT GTA CCT TGC TGC TAT GAA AAC AAC GCA GAG GCC Phe Gln Pro Pro Cys Val Pro Cys Cys Tyr Glu Asn Asn Ala Glu Ala 290 295 300	8323
ACA CTA CGG ATG CTC GAG GAT AAC GTG GAT AGG CCA GGG TAC TAC GAC Thr Leu Arg Met Leu Glu Asp Asn Val Asp Arg Pro Gly Tyr Tyr Asp 305 310 315	8371
CTC CTT CAG GCA GCC TTG ACG TGC CGA AAC GGA ACA AGA CAC CGG CGC Leu Leu Gln Ala Ala Leu Thr Cys Arg Asn Gly Thr Arg His Arg Arg 320 325 330	8419

AGC GTG TCG CAA CAC TTC AAC GTG TAT AAG GCT ACA CGC CCT TAC ATC Ser Val Ser Gln His Phe Asn Val Tyr Lys Ala Thr Arg Pro Tyr Ile 335 340 345	8467
GCG TAC TGC GCC GAC TGC GGA GCA GGG CAC TCG TGT CAT AGC CCC GTA Ala Tyr Cys Ala Asp Cys Gly Ala Gly His Ser Cys His Ser Pro Val 350 355 360 365	8515
GCA ATT GAA GCG GTC AGG TCC GAA GCT ACC GAC GGG ATG CTG AAG ATT Ala Ile Glu Ala Val Arg Ser Glu Ala Thr Asp Gly Met Leu Lys Ile 370 375 380	8563
CAG TTC TCG GCA CAA ATT GGC ATA GAT AAG AGT GAC AAT CAT GAC TAC Gln Phe Ser Ala Gln Ile Gly Ile Asp Lys Ser Asp Asn His Asp Tyr 385 390 395	8611
ACG AAG ATA AGG TAC GCA GAC GGG CAC GCC ATT GAG AAT GCC GTC CGG Thr Lys Ile Arg Tyr Ala Asp Gly His Ala Ile Glu Asn Ala Val Arg 400 405 410	8659
TCA TCT TTG AAG GTA GCC ACC TCC GGA GAC TGT TTC GTC CAT GGC ACA Ser Ser Leu Lys Val Ala Thr Ser Gly Asp Cys Phe Val His Gly Thr 415 420 425	8707
ATG GGA CAT TTC ATA CTG GCA AAG TGC CCA CCG GGT GAA TTC CTG CAG Met Gly His Phe Ile Leu Ala Lys Cys Pro Pro Gly Glu Phe Leu Gln 430 435 440 445	8755
GTC TCG ATC CAG GAC ACC AGA AAC GCG GTC CGT GCC TGC AGA ATA CAA Val Ser Ile Gln Asp Thr Arg Asn Ala Val Arg Ala Cys Arg Ile Gln 450 455 460	8803
TAT CAT CAT GAC CCT CAA CCG GTG GGT AGA GAA AAA TTT ACA ATT AGA Tyr His His Asp Pro Gln Pro Val Gly Arg Glu Lys Phe Thr Ile Arg 465 470 475	8851
CCA CAC TAT GGA AAA GAG ATC CCT TGC ACC ACT TAT CAA CAG ACC ACA Pro His Tyr Gly Lys Glu Ile Pro Cys Thr Thr Tyr Gln Gln Thr Thr 480 485 490	8899
GCG AAG ACC GTG GAG GAA ATC GAC ATG CAT ATG CCG CCA GAT ACG CCG Ala Lys Thr Val Glu Glu Ile Asp Met His Met Pro Pro Asp Thr Pro 495 500 505	8947
GAC AGG ACG TTG CTA TCA CAG CAA TCT GGC AAT GTA AAG ATC ACA GTC Asp Arg Thr Leu Leu Ser Gln Gln Ser Gly Asn Val Lys Ile Thr Val 510 515 520 525	8995
GGA GGA AAG AAG GTG AAA TAC AAC TGC ACC TGT GGA ACC GGA AAC GTT Gly Gly Lys Lys Val Lys Tyr Asn Cys Thr Cys Gly Thr Gly Asn Val 530 535 540	9043

GGC ACT ACT AAT TCG GAC ATG ACG ATC AAC ACG TGT CTA ATA GAG CAG Gly Thr Thr Asn Ser Asp Met Thr Ile Asn Thr Cys Leu Ile Glu Gln 545 550 555	9091
TGC CAC GTC TCA GTG ACG GAC CAT AAG AAA TGG CAG TTC AAC TCA CCT Cys His Val Ser Val Thr Asp His Lys Lys Trp Gln Phe Asn Ser Pro 560 565 570	9139
TTC GTC CCG AGA GCC GAC GAA CCG GCT AGA AAA GGC AAA GTC CAT ATC Phe Val Pro Arg Ala Asp Glu Pro Ala Arg Lys Gly Lys Val His Ile 575 580 585	9187
CCA TTC CCG TTG GAC AAC ATC ACA TGC AGA GTT CCA ATG GCG CGC GAA Pro Phe Pro Leu Asp Asn Ile Thr Cys Arg Val Pro Met Ala Arg Glu 590 595 600 605	9235
CCA ACC GTC ATC CAC GGC AAA AGA GAA GTG ACA CTG CAC CTT CAC CCA Pro Thr Val Ile His Gly Lys Arg Glu Val Thr Leu His Leu His Pro 610 615 620	9283
GAT CAT CCC ACG CTC TTT TCC TAC CGC ACA CTG GGT GAG GAC CCG CAG Asp His Pro Thr Leu Phe Ser Tyr Arg Thr Leu Gly Glu Asp Pro Gln 625 630 635	9331
TAT CAC GAG GAA TGG GTG ACA GCG GCG GTG GAA CGG ACC ATA CCC GTA Tyr His Glu Glu Trp Val Thr Ala Ala Val Glu Arg Thr Ile Pro Val 640 645 650	9379
CCA GTG GAC GGG ATG GAG TAC CAC TGG GGA AAC AAC GAC CCA GTG AGG Pro Val Asp Gly Met Glu Tyr His Trp Gly Asn Asn Asp Pro Val Arg 655 660 665	9427
CTT TGG TCT CAA CTC ACC ACT GAA GGG AAA CCG CAC GGC TGG CCG CAT Leu Trp Ser Gln Leu Thr Thr Glu Gly Lys Pro His Gly Trp Pro His 670 675 680 685	9475
CAG ATC GTA CAG TAC TAC TAT GGG CTT TAC CCG GCC GCT ACA GTA TCC Gln Ile Val Gln Tyr Tyr Tyr Gly Leu Tyr Pro Ala Ala Thr Val Ser 690 695 700	9523
GCG GTC GTC GGG ATG AGC TTA CTG GCG TTG ATA TCG ATC TTC GCG TCG Ala Val Val Gly Met Ser Leu Leu Ala Leu Ile Ser Ile Phe Ala Ser 705 710 715	9571
TGC TAC ATG CTG GTT GCG GCC CGC AGT AAG TGC TTG ACC CCT TAT GCT Cys Tyr Met Leu Val Ala Ala Arg Ser Lys Cys Leu Thr Pro Tyr Ala 720 725 730	9619
TTA ACA CCA GGA GCT GCA GTT CCG TGG ACG CTG GGG ATA CTC TGC TGC Leu Thr Pro Gly Ala Ala Val Pro Trp Thr Leu Gly Ile Leu Cys Cys 735 740 745	9667

GCC CCG CGG GCG CAC GCA GCT AGT GTG GCA GAG ACT ATG GCC TAC TTG Ala Pro Arg Ala His Ala Ala Ser Val Ala Glu Thr Met Ala Tyr Leu 750 755 760 765	9715
TGG GAC CAA AAC CAA GCG TTG TTC TGG TTG GAG TTT GCG GCC CCT GTT Trp Asp Gln Asn Gln Ala Leu Phe Trp Leu Glu Phe Ala Ala Pro Val 770 775 780	9763
GCC TGC ATC CTC ATC ATC ACG TAT TGC CTC AGA AAC GTG CTG TGT TGC Ala Cys Ile Leu Ile Ile Thr Tyr Cys Leu Arg Asn Val Leu Cys Cys 785 790 795	9811
TGT AAG AGC CTT TCT TTT TTA GTG CTA CTG AGC CTC GGG GCA ACC GCC Cys Lys Ser Leu Ser Phe Leu Val Leu Leu Ser Leu Gly Ala Thr Ala 800 805 810	9859
AGA GCT TAC GAA CAT TCG ACA GTA ATG CCG AAC GTG GTG GGG TTC CCG Arg Ala Tyr Glu His Ser Thr Val Met Pro Asn Val Val Gly Phe Pro 815 820 825	9907
TAT AAG GCT CAC ATT GAA AGG CCA GGA TAT AGC CCC CTC ACT TTG CAG Tyr Lys Ala His Ile Glu Arg Pro Gly Tyr Ser Pro Leu Thr Leu Gln 830 835 840 845	9955
ATG CAG GTT GAA ACC AGC CTC GAA CCA ACC CTT AAT TTG GAA TAC Met Gln Val Val Glu Thr Ser Leu Glu Pro Thr Leu Asn Leu Glu Tyr 850 855 860	10003
ATA ACC TGT GAG TAC AAG ACG GTC GTC CCG TCG CCG TAC GTG AAG TGC Ile Thr Cys Glu Tyr Lys Thr Val Val Pro Ser Pro Tyr Val Lys Cys 865 870 875	10051
TGC GGC GCC TCA GAG TGC TCC ACT AAA GAG AAG CCT GAC TAC CAA TGC Cys Gly Ala Ser Glu Cys Ser Thr Lys Glu Lys Pro Asp Tyr Gln Cys 880 885 890	10099
AAG GTT TAC ACA GGC GTG TAC CCG TTC ATG TGG GGA GGG GCA TAT TGC Lys Val Tyr Thr Gly Val Tyr Pro Phe Met Trp Gly Gly Ala Tyr Cys 895 900 905	10147
TTC TGC GAC TCA GAA AAC ACG CAA CTC AGC GAG GCG TAC GTC GAT CGA Phe Cys Asp Ser Glu Asn Thr Gln Leu Ser Glu Ala Tyr Val Asp Arg 910 915 920 925	10195
TCG GAC GTA TGC AGG CAT GAT CAC GCA TCT GCT TAC AAA GCC CAT ACA Ser Asp Val Cys Arg His Asp His Ala Ser Ala Tyr Lys Ala His Thr 930 935 940	10243
GCA TCG CTG AAG GCC AAA GTG AGG GTT ATG TAC GGC AAC GTA AAC CAG Ala Ser Leu Lys Ala Lys Val Arg Val Met Tyr Gly Asn Val Asn Gln 945 950 955	10291

ACT GTG GAT GTT TAC GTG AAC GGA GAC CAT GCC GTC ACG ATA GGG GGT		10339
Thr Val Asp Val Tyr Val Asn Gly Asp His Ala Val Thr Ile Gly Gly		
960 965 970		
ACT CAG TTC ATA TTC GGG CCG CTG TCA TCG GCC TGG ACC CCG TTC GAC		10387
Thr Gln Phe Ile Phe Gly Pro Leu Ser Ser Ala Trp Thr Pro Phe Asp		
975 980 985		
AAC AAG ATA GTC GTG TAC AAA GAC GAA GTG TTC AAT CAG GAC TTC CCG		10435
Asn Lys Ile Val Val Tyr Lys Asp Glu Val Phe Asn Gln Asp Phe Pro		
990 995 1000 1005		
CCG TAC GGA TCT GGG CAA CCA GGG CGC TTC GGC GAC ATC CAA AGC AGA		10483
Pro Tyr Gly Ser Gly Gln Pro Gly Arg Phe Gly Asp Ile Gln Ser Arg		
1010 1015 1020		
ACA GTG GAG AGT AAC GAC CTG TAC GCG AAC ACG GCA CTG AAG CTG GCA		10531
Thr Val Glu Ser Asn Asp Leu Tyr Ala Asn Thr Ala Leu Lys Leu Ala		
1025 1030 1035		
CGC CCT TCA CCC GGC ATG GTC CAT GTA CCG TAC ACA CAG ACA CCT TCA		10579
Arg Pro Ser Pro Gly Met Val His Val Pro Tyr Thr Gln Thr Pro Ser		
1040 1045 1050		
GGG TTC AAA TAT TGG CTA AAG GAA AAA GGG ACA GCC CTA AAT ACG AAG		10627
Gly Phe Lys Tyr Trp Leu Lys Glu Lys Gly Thr Ala Leu Asn Thr Lys		
1055 1060 1065		
GCT CCT TTT GGC TGC CAA ATC AAA ACG AAC CCT GTC AGG GCC ATG AAC		10675
Ala Pro Phe Gly Cys Gln Ile Lys Thr Asn Pro Val Arg Ala Met Asn		
1070 1075 1080 1085		
TGC GCC GTG GGA AAC ATC CCT GTC TCC ATG AAT TTG CCT GAC AGC GCC		10723
Cys Ala Val Gly Asn Ile Pro Val Ser Met Asn Leu Pro Asp Ser Ala		
1090 1095 1100		
TTT ACC CGC ATT GTC GAG GCG CCG ACC ATC ATT GAC CTG ACT TGC ACA		10771
Phe Thr Arg Ile Val Glu Ala Pro Thr Ile Ile Asp Leu Thr Cys Thr		
1105 1110 1115		
GTG GCT ACC TGT ACG CAC TCC TCG GAT TTC GGC GGC GTC TTG ACA CTG		10819
Val Ala Thr Cys Thr His Ser Ser Asp Phe Gly Gly Val Leu Thr Leu		
1120 1125 1130		
ACG TAC AAG ACC AAC AAG AAC GGG GAC TGC TCT GTA CAC TCG CAC TCT		10867
Thr Tyr Lys Thr Asn Lys Asn Gly Asp Cys Ser Val His Ser His Ser		
1135 1140 1145		
AAC GTA GCT ACT CTA CAG GAG GCC ACA GCA AAA GTG AAG ACA GCA GGT		10915
Asn Val Ala Thr Leu Gln Glu Ala Thr Ala Lys Val Lys Thr Ala Gly		
1150 1155 1160 1165		

AAG GTG ACC TTA CAC TTC TCC ACG GCA AGC GCA TCA CCT TCT TTT GTG Lys Val Thr Leu His Phe Ser Thr Ala Ser Ala Ser Pro Ser Phe Val 1170 1175 1180	10963
GTG TCG CTA TGC AGT GCT AGG GCC ACC TGT TCA GCG TCG TGT GAG CCC Val Ser Leu Cys Ser Ala Arg Ala Thr Cys Ser Ala Ser Cys Glu Pro 1185 1190 1195	11011
CCG AAA GAC CAC ATA GTC CCA TAT GCG GCT AGC CAC AGT AAC GTA GTG Pro Lys Asp His Ile Val Pro Tyr Ala Ala Ser His Ser Asn Val Val 1200 1205 1210	11059
TTT CCA GAC ATG TCG GGC ACC GCA CTA TCA TGG GTG CAG AAA ATC TCG Phe Pro Asp Met Ser Gly Thr Ala Leu Ser Trp Val Gln Lys Ile Ser 1215 1220 1225	11107
GGT GGT CTG GGG GCC TTC GCA ATC GGC GCT ATC CTG GTG CTG GTT GTG Gly Gly Leu Gly Ala Phe Ala Ile Gly Ala Ile Leu Val Leu Val Val 1230 1235 1240 1245	11155
GTC ACT TGC ATT GGG CTC CGC AGA TAAGTTAGGG TAGGCAATGG CATTGATATA Val Thr Cys Ile Gly Leu Arg Arg 1250	11209
GCAAGAAAAT TGAAAACAGA AAAAGTTAGG GTAAGCAATG GCATATAACC ATAACGTAT 1269	11269
AACTTGTAAC AAAGCGAAC AAGACCTGCG CAATTGGCCC CGTGGTCCGC CTCACGGAAA 1329	11329
CTCGGGGCAA CTCATATTGA CACATTAATT GGCAATAATT GGAAGCTTAC ATAAGCTTAA 1389	11389
TCGACGAAT AATTGGATT TTATTTATT TTGCAATTGG TTTTAATAT TTCCAAAAAA 1449	11449
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1509	11509
AAAACTAG	11517

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2431 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Ala Lys Val His Val Asp Ile Glu Ala Asp Ser Pro Phe Ile 1 5 10 15
--

Lys Ser Leu Gln Lys Ala Phe Pro Ser Phe Glu Val Glu Ser Leu Gln

20

25

30

Val Thr Pro Asn Asp His Ala Asn Ala Arg Ala Phe Ser His Leu Ala
 35 40 45

Thr Lys Leu Ile Glu Gln Glu Thr Asp Lys Asp Thr Leu Ile Leu Asp
 50 55 60

Ile Gly Ser Ala Pro Ser Arg Arg Met Met Ser Thr His Lys Tyr His
 65 70 75 80

Cys Val Cys Pro Met Arg Ser Ala Glu Asp Pro Glu Arg Leu Asp Ser
 85 90 95

Tyr Ala Lys Lys Leu Ala Ala Ser Gly Lys Val Leu Asp Arg Glu
 100 105 110

Ile Ala Gly Lys Ile Thr Asp Leu Gln Thr Val Met Ala Thr Pro Asp
 115 120 125

Ala Glu Ser Pro Thr Phe Cys Leu His Thr Asp Val Thr Cys Arg Thr
 130 135 140

Ala Ala Glu Val Ala Val Tyr Gln Asp Val Tyr Ala Val His Ala Pro
 145 150 155 160

Thr Ser Leu Tyr His Gln Ala Met Lys Gly Val Arg Thr Ala Tyr Trp
 165 170 175

Ile Gly Phe Asp Thr Thr Pro Phe Met Phe Asp Ala Leu Ala Gly Ala
 180 185 190

Tyr Pro Thr Tyr Ala Thr Asn Trp Ala Asp Glu Gln Val Leu Gln Ala
 195 200 205

Arg Asn Ile Gly Leu Cys Ala Ala Ser Leu Thr Glu Gly Arg Leu Gly
 210 215 220

Lys Leu Ser Ile Leu Arg Lys Lys Gln Leu Lys Pro Cys Asp Thr Val
 225 230 235 240

Met Phe Ser Val Gly Ser Thr Leu Tyr Thr Glu Ser Arg Lys Leu Leu
 245 250 255

Arg Ser Trp His Leu Pro Ser Val Phe His Leu Lys Gly Lys Gln Ser
 260 265 270

Phe Thr Cys Arg Cys Asp Thr Ile Val Ser Cys Glu Gly Tyr Val Val
 275 280 285

Lys Lys Ile Thr Met Cys Pro Gly Leu Tyr Gly Lys Thr Val Gly Tyr
 290 295 300

Ala Val Thr Tyr His Ala Glu Gly Phe Leu Val Cys Lys Thr Thr Asp
 305 310 315 320
 Thr Val Lys Gly Glu Arg Val Ser Phe Pro Val Cys Thr Tyr Val Pro
 325 330 335
 Ser Thr Ile Cys Asp Gln Met Thr Gly Ile Leu Ala Thr Asp Val Thr
 340 345 350
 Pro Glu Asp Ala Gln Lys Leu Leu Val Gly Leu Asn Gln Arg Ile Val
 355 360 365
 Val Asn Gly Arg Thr Gln Arg Asn Thr Asn Thr Met Lys Asn Tyr Leu
 370 375 380
 Leu Pro Ile Val Ala Val Ala Phe Ser Lys Trp Ala Arg Glu Tyr Lys
 385 390 395 400
~~Ala~~ Asp Leu Asp Asp Glu Lys Pro Leu Gly Val Arg Glu Arg Ser Leu
 405 410 415
~~Thr~~ Cys Cys Cys Leu Trp Ala Phe Lys Thr Arg Lys Met His Thr Met
 420 425 430
~~Tyr~~ Lys Lys Pro Asp Thr Gln Thr Ile Val Lys Val Pro Ser Glu Phe
 435 440 445
~~Asn~~ Ser Phe Val Ile Pro Ser Leu Trp Ser Thr Gly Leu Ala Ile Pro
 450 455 460
~~Val~~ Arg Ser Arg Ile Lys Met Leu Leu Ala Lys Lys Thr Lys Arg Glu
 465 470 475 480
~~Ile~~ Ile Pro Val Leu Asp Ala Ser Ser Ala Arg Asp Ala Glu Gln Glu
 485 490 495
 Glu Lys Glu Arg Leu Glu Ala Glu Leu Thr Arg Glu Ala Leu Pro Pro
 500 505 510
 Leu Val Pro Ile Ala Pro Ala Glu Thr Gly Val Val Asp Val Asp Val
 515 520 525
 Glu Glu Leu Glu Tyr His Ala Gly Ala Gly Val Val Glu Thr Pro Arg
 530 535 540
 Ser Ala Leu Lys Val Thr Ala Gln Pro Asn Asp Val Leu Leu Gly Asn
 545 550 555 560
 Tyr Val Val Leu Ser Pro Gln Thr Val Leu Lys Ser Ser Lys Leu Ala
 565 570 575

Pro Val His Pro Leu Ala Glu Gln Val Lys Ile Ile Thr His Asn Gly
 580 585 590

Arg Ala Gly Gly Tyr Gln Val Asp Gly Tyr Asp Gly Arg Val Leu Leu
 595 600 605

Pro Cys Gly Ser Ala Ile Pro Val Pro Glu Phe Gln Ala Leu Ser Glu
 610 615 620

Ser Ala Thr Met Val Tyr Asn Glu Arg Glu Phe Val Asn Arg Lys Leu
 625 630 635 640

Tyr His Ile Ala Val His Gly Pro Ser Leu Asn Thr Asp Glu Glu Asn
 645 650 655

Tyr Glu Lys Val Arg Ala Glu Arg Thr Asp Ala Glu Tyr Val Phe Asp
 660 665 670

Val Asp Lys Lys Cys Cys Val Lys Arg Glu Glu Ala Ser Gly Leu Val
 675 680 685

Leu Val Gly Glu Leu Thr Asn Pro Pro Phe His Glu Phe Ala Tyr Glu
 690 695 700

Gly Leu Lys Ile Arg Pro Ser Ala Pro Tyr Lys Thr Thr Val Val Gly
 705 710 715 720

Val Phe Gly Val Pro Gly Ser Gly Lys Ser Ala Ile Ile Lys Ser Leu
 725 730 735

Val Thr Lys His Asp Leu Val Thr Ser Gly Lys Lys Glu Asn Cys Gln
 740 745 750

Glu Ile Val Asn Asp Val Lys Lys His Arg Gly Lys Gly Thr Ser Arg
 755 760 765

Glu Asn Ser Asp Ser Ile Leu Leu Asn Gly Cys Arg Arg Ala Val Asp
 770 775 780

Ile Leu Tyr Val Asp Glu Ala Phe Ala Cys His Ser Gly Thr Leu Leu
 785 790 795 800

Ala Leu Ile Ala Leu Val Lys Pro Arg Ser Lys Val Val Leu Cys Gly
 805 810 815

Asp Pro Lys Gln Cys Gly Phe Phe Asn Met Met Gln Leu Lys Val Asn
 820 825 830

Phe Asn His Asn Ile Cys Thr Glu Val Cys His Lys Ser Ile Ser Arg
 835 840 845

Arg Cys Thr Arg Pro Val Thr Ala Ile Val Ser Thr Leu His Tyr Gly
 850 855 860

Gly Lys Met Arg Thr Thr Asn Pro Cys Asn Lys Pro Ile Ile Ile Asp
 865 870 875 880

Thr Thr Gly Gln Thr Lys Pro Lys Pro Gly Asp Ile Val Leu Thr Cys
 885 890 895

Phe Arg Gly Trp Ala Lys Gln Leu Gln Leu Asp Tyr Arg Gly His Glu
 900 905 910

Val Met Thr Ala Ala Ala Ser Gln Gly Leu Thr Arg Lys Gly Val Tyr
 915 920 925

Ala Val Arg Gln Lys Val Asn Glu Asn Pro Leu Tyr Ala Pro Ala Ser
 930 935 940

His Val Asn Val Leu Leu Thr Arg Thr Glu Asp Arg Leu Val Trp
 945 950 955 960

Lys Thr Leu Ala Gly Asp Pro Trp Ile Lys Val Leu Ser Asn Ile Pro
 965 970 975

Gly Asn Phe Thr Ala Thr Leu Glu Glu Trp Gln Glu Glu His Asp
 980 985 990

Lys Ile Met Lys Val Ile Glu Gly Pro Ala Ala Pro Val Asp Ala Phe
 995 1000 1005

Asn Lys Ala Asn Val Cys Trp Ala Lys Ser Leu Val Pro Val Leu
 1010 1015 1020

Asp Thr Ala Gly Ile Arg Leu Thr Ala Glu Glu Trp Ser Thr Ile Ile
 1025 1030 1035 1040

Thr Ala Phe Lys Glu Asp Arg Ala Tyr Ser Pro Val Val Ala Leu Asn
 1045 1050 1055

Glu Ile Cys Thr Lys Tyr Tyr Gly Val Asp Leu Asp Ser Gly Leu Phe
 1060 1065 1070

Ser Ala Pro Lys Val Ser Leu Tyr Tyr Glu Asn Asn His Trp Asp Asn
 1075 1080 1085

Arg Pro Gly Gly Arg Met Tyr Gly Phe Asn Ala Ala Thr Ala Ala Arg
 1090 1095 1100

Leu Glu Ala Arg His Thr Phe Leu Lys Gly Gln Trp His Thr Gly Lys
 1105 1110 1115 1120

Asn Val Ile Pro Ile Asn Arg Arg Leu Pro His Ala Leu Val Ala Glu
1140 1145 1150

Tyr Lys Thr Val Lys Gly Ser Arg Val Glu Trp Leu Val Asn Lys Val
1155 1160 1165

Arg Gly Tyr His Val Leu Leu Val Ser Glu Tyr Asn Leu Ala Leu Pro
1170 1175 1180

Arg Arg Arg Val Thr Trp Leu Ser Pro Leu Asn Val Thr Gly Ala Asp
 1185 1190 1195 1200

Arg Cys Tyr Asp Leu Ser Leu Gly Leu Pro Ala Asp Ala Gly Arg Phe
1205 1210 1215

Asp Leu Val Phe Val Asn Ile His Thr Glu Phe Arg Ile His His Tyr
1220 1225 1230

Gin Gln Cys Val Asp His Ala Met Lys Leu Gln Met Leu Gly Gly Asp
H 1235 1240 1245

A 1250 **G** 1255 **A** 1260
Leu Arg Leu Leu Lys Pro Gly Gly Ile Leu Met Arg Ala Tyr Gly

Tyr Ala Asp Lys Ile Ser Glu Ala Val Val Ser Ser Leu Ser Arg Lys
1265 1270 1275 128

Dle Ser Ser Ala Arg Val Leu Arg Pro Asp Cys Val Thr Ser Asn Thr
□ 1285 1290 1295

Glu Val Phe Leu Leu Phe Ser Asn Phe Asp Asn Gly Lys Arg Pro Ser
1300 1305 1310

Thr Leu His Gln Met Asn Thr Lys Leu Ser Ala Val Tyr Ala Gly Glu
 1315 1320 1325

Ala Met His Thr Ala Gly Cys Ala Pro Ser Tyr Arg Val Lys Arg Ala
1330 1335 1340

Asp Ile Ala Thr Cys Thr Glu Ala Ala Val Val Asn Ala Ala Asn Ala
1345 1350 1355 1360

Arg Gly Thr Val Gly Asp Gly Val Cys Arg Ala Val Ala Lys Lys Trp
1365 1370 1375

Pro Ser Ala Phe Lys Gly Ala Ala Thr Pro Val Gly Thr Ile Lys Thr
1380 1385 1390

Val Met Cys Gly Ser Tyr Pro Val Ile His Ala Val Ala Pro Asn Phe

1395

1400

1405

Ser Ala Thr Thr Glu Ala Glu Gly Asp Arg Glu Leu Ala Ala Val Tyr
 1410 1415 1420

Arg Ala Val Ala Ala Glu Val Asn Arg Leu Ser Leu Ser Ser Val Ala
 1425 1430 1435 1440

Ile Pro Leu Leu Ser Thr Gly Val Phe Ser Gly Gly Arg Asp Arg Leu
 1445 1450 1455

Gln Gln Ser Leu Asn His Leu Phe Thr Ala Met Asp Ala Thr Asp Ala
 1460 1465 1470

Asp Val Thr Ile Tyr Cys Arg Asp Lys Ser Trp Glu Lys Lys Ile Gln
 1475 1480 1485

Glu Ala Ile Asp Met Arg Thr Ala Val Glu Leu Leu Asn Asp Asp Val
 1490 1495 1500

Glu Leu Thr Thr Asp Leu Val Arg Val His Pro Asp Ser Ser Leu Val
 1505 1510 1515 1520

Gly Arg Lys Gly Tyr Ser Thr Thr Asp Gly Ser Leu Tyr Ser Tyr Phe
 1525 1530 1535

Glu Gly Thr Lys Phe Asn Gln Ala Ala Ile Asp Met Ala Glu Ile Leu
 1540 1545 1550

Thr Leu Trp Pro Arg Leu Gln Glu Ala Asn Glu Gln Ile Cys Leu Tyr
 1555 1560 1565

Ala Leu Gly Glu Thr Met Asp Asn Ile Arg Ser Lys Cys Pro Val Asn
 1570 1575 1580

Asp Ser Asp Ser Ser Thr Pro Pro Arg Thr Val Pro Cys Leu Cys Arg
 1585 1590 1595 1600

Tyr Ala Met Thr Ala Glu Arg Ile Ala Arg Leu Arg Ser His Gln Val
 1605 1610 1615

Lys Ser Met Val Val Cys Ser Ser Phe Pro Leu Pro Lys Tyr His Val
 1620 1625 1630

Asp Gly Val Gln Lys Val Lys Cys Glu Lys Val Leu Leu Phe Asp Pro
 1635 1640 1645

Thr Val Pro Ser Val Val Ser Pro Arg Lys Tyr Ala Ala Ser Thr Thr
 1650 1655 1660

Asp His Ser Asp Arg Ser Leu Arg Gly Phe Asp Leu Asp Trp Thr Thr
 1665 1670 1675 1680

Asp Ser Ser Ser Thr Ala Ser Asp Thr Met Ser Leu Pro Ser Leu Gln
 1685 1690 1695

Ser Cys Asp Ile Asp Ser Ile Tyr Glu Pro Met Ala Pro Ile Val Val
 1700 1705 1710

Thr Ala Asp Val His Pro Glu Pro Ala Gly Ile Ala Asp Leu Ala Ala
 1715 1720 1725

Asp Val His Pro Glu Pro Ala Asp His Val Asp Leu Glu Asn Pro Ile
 1730 1735 1740

Pro Pro Pro Arg Pro Lys Arg Ala Ala Tyr Leu Ala Ser Arg Ala Ala
 1745 1750 1755 1760

Glu Arg Pro Val Pro Ala Pro Arg Lys Pro Thr Pro Ala Pro Arg Thr
 1765 1770 1775

~~A~~ Phe Arg Asn Lys Leu Pro Leu Thr Phe Gly Asp Phe Asp Glu His
 1780 1785 1790

~~G~~lu Val Asp Ala Leu Ala Ser Gly Ile Thr Phe Gly Asp Phe Asp Asp
 1795 1800 1805

~~V~~al Leu Arg Leu Gly Arg Ala Gly Ala Tyr Ile Phe Ser Ser Asp Thr
 1810 1815 1820

~~G~~ly Ser Gly His Leu Gln Gln Lys Ser Val Arg Gln His Asn Leu Gln
 1825 1830 1835 1840

~~C~~ys Ala Gln Leu Asp Ala Val Gln Glu Lys Met Tyr Pro Pro Lys
 1845 1850 1855

~~L~~eu Asp Thr Glu Arg Glu Lys Leu Leu Leu Lys Met Gln Met His
 1860 1865 1870

Pro Ser Glu Ala Asn Lys Ser Arg Tyr Gln Ser Arg Lys Val Glu Asn
 1875 1880 1885

Met Lys Ala Thr Val Val Asp Arg Leu Thr Ser Gly Ala Arg Leu Tyr
 1890 1895 1900

Thr Gly Ala Asp Val Gly Arg Ile Pro Thr Tyr Ala Val Arg Tyr Pro
 1905 1910 1915 1920

Arg Pro Val Tyr Ser Pro Thr Val Ile Glu Arg Phe Ser Ser Pro Asp
 1925 1930 1935

Val Ala Ile Ala Ala Cys Asn Glu Tyr Leu Ser Arg Asn Tyr Pro Thr
 1940 1945 1950

Val Ala Ser Tyr Gln Ile Thr Asp Glu Tyr Asp Ala Tyr Leu Asp Met
 1955 1960 1965

Val Asp Gly Ser Asp Ser Cys Leu Asp Arg Ala Thr Phe Cys Pro Ala
 1970 1975 1980

Lys Leu Arg Cys Tyr Pro Lys His His Ala Tyr His Gln Pro Thr Val
 1985 1990 1995 2000

Arg Ser Ala Val Pro Ser Pro Phe Gln Asn Thr Leu Gln Asn Val Leu
 2005 2010 2015

Ala Ala Ala Thr Lys Arg Asn Cys Asn Val Thr Gln Met Arg Glu Leu
 2020 2025 2030

Pro Thr Met Asp Ser Ala Val Phe Asn Val Glu Cys Phe Lys Arg Tyr
 2035 2040 2045

~~A~~ Cys Ser Gly Glu Tyr Trp Glu Glu Tyr Ala Lys Gln Pro Ile Arg
 2050 2055 2060

~~I~~le Thr Thr Glu Asn Ile Thr Tyr Val Thr Lys Leu Lys Gly Pro
 2065 2070 2075 2080

~~L~~ys Ala Ala Ala Leu Phe Ala Lys Thr His Asn Leu Val Pro Leu Gln
 2085 2090 2095

~~G~~lu Val Pro Met Asp Arg Phe Thr Val Asp Met Lys Arg Asp Val Lys
 2100 2105 2110

~~V~~al Thr Pro Gly Thr Lys His Thr Glu Glu Arg Pro Lys Val Gln Val
 2115 2120 2125

~~I~~le Gln Ala Ala Glu Pro Leu Ala Thr Ala Tyr Leu Cys Gly Ile His
 2130 2135 2140

Arg Glu Leu Val Arg Arg Leu Asn Ala Val Leu Arg Pro Asn Val His
 2145 2150 2155 2160

Thr Leu Phe Asp Met Ser Ala Glu Asp Phe Asp Ala Ile Ile Ala Ser
 2165 2170 2175

His Phe His Pro Gly Asp Pro Val Leu Glu Thr Asp Ile Ala Ser Phe
 2180 2185 2190

Asp Lys Ser Gln Asp Asp Ser Leu Ala Leu Thr Gly Leu Met Ile Leu
 2195 2200 2205

Glu Asp Leu Gly Val Asp Gln Tyr Leu Leu Asp Leu Ile Glu Ala Ala
 2210 2215 2220

Phe Gly Glu Ile Ser Ser Cys His Leu Pro Thr Gly Thr Arg Phe Lys
 2225 2230 2235 2240

 Phe Gly Ala Met Met Lys Ser Gly Met Phe Leu Thr Leu Phe Ile Asn
 2245 2250 2255

 Thr Val Leu Asn Ile Thr Ile Ala Ser Arg Val Leu Glu Gln Arg Leu
 2260 2265 2270

 Thr Asp Ser Ala Cys Ala Ala Phe Ile Gly Asp Asp Asn Ile Val His
 2275 2280 2285

 Gly Val Ile Ser Asp Lys Leu Met Ala Glu Arg Cys Ala Ser Trp Val
 2290 2295 2300

 Asn Met Glu Val Lys Ile Ile Asp Ala Val Met Gly Glu Lys Pro Pro
 2305 2310 2315 2320

 Tyr Phe Cys Gly Gly Phe Ile Val Phe Asp Ser Val Thr Gln Thr Ala
 2325 2330 2335

 Cys Arg Val Ser Asp Pro Leu Lys Arg Leu Phe Lys Leu Gly Lys Pro
 2340 2345 2350

 Leu Thr Ala Glu Asp Lys Gln Asp Glu Asp Arg Arg Ala Leu Ser
 2355 2360 2365

 Asp Glu Val Ser Lys Trp Phe Arg Thr Gly Leu Gly Ala Glu Leu Glu
 2370 2375 2380

 Val Ala Leu Thr Ser Arg Tyr Glu Val Glu Gly Cys Lys Ser Ile Leu
 2385 2390 2395 2400

 Ile Ala Met Thr Thr Leu Ala Arg Asp Ile Lys Ala Phe Lys Lys Leu
 2405 2410 2415

 Arg Gly Pro Val Ile His Leu Tyr Gly Gly Pro Arg Leu Val Arg
 2420 2425 2430

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1253 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met	Asn	Tyr	Ile	Pro	Thr	Gln	Thr	Phe	Tyr	Gly	Arg	Arg	Trp	Arg	Pro
1									10					15	
Arg	Pro	Ala	Ala	Arg	Pro	Trp	Pro	Leu	Gln	Ala	Thr	Pro	Val	Ala	Pro
								25					30		
Val	Val	Pro	Asp	Phe	Gln	Ala	Gln	Gln	Met	Gln	Gln	Leu	Ile	Ser	Ala
							35	40				45			
Val	Asn	Ala	Leu	Thr	Met	Arg	Gln	Asn	Ala	Ile	Ala	Pro	Ala	Arg	Pro
					50	55				60					
Pro	Lys	Pro	Lys	Lys	Lys	Lys	Thr	Thr	Lys	Pro	Lys	Pro	Lys	Thr	Gln
	65						70		75			80			
Pro	Lys	Lys	Ile	Asn	Gly	Lys	Thr	Gln	Gln	Lys	Lys	Lys	Asp	Lys	
						85			90			95			
Gln	Ala	Asp	Lys	Lys	Lys	Lys	Pro	Gly	Lys	Arg	Glu	Arg	Met	Cys	
						100		105			110				
Met	Lys	Ile	Glu	Asn	Asp	Cys	Ile	Phe	Glu	Val	Lys	His	Glu	Gly	Lys
						115		120			125				
Val	Thr	Gly	Tyr	Ala	Cys	Leu	Val	Gly	Asp	Lys	Val	Met	Lys	Pro	Ala
						130		135			140				
His	Val	Lys	Gly	Val	Ile	Asp	Asn	Ala	Asp	Leu	Ala	Lys	Leu	Ala	Phe
	45					150				155			160		
Lys	Lys	Ser	Ser	Lys	Tyr	Asp	Leu	Glu	Cys	Ala	Gln	Ile	Pro	Val	His
						165			170			175			
Met	Arg	Ser	Asp	Ala	Ser	Lys	Tyr	Thr	His	Glu	Lys	Pro	Glu	Gly	His
						180		185			190				
Tyr	Asn	Trp	His	His	Gly	Ala	Val	Gln	Tyr	Ser	Gly	Gly	Arg	Phe	Thr
							195		200			205			
Ile	Pro	Thr	Gly	Ala	Gly	Lys	Pro	Gly	Asp	Ser	Gly	Arg	Pro	Ile	Phe
						210		215			220				
Asp	Asn	Lys	Gly	Arg	Val	Val	Ala	Ile	Val	Leu	Gly	Gly	Ala	Asn	Glu
						225		230			235			240	
Gly	Ser	Arg	Thr	Ala	Leu	Ser	Val	Val	Thr	Trp	Asn	Lys	Asp	Met	Val
						245			250			255			
Thr	Arg	Val	Thr	Pro	Glu	Gly	Ser	Glu	Glu	Trp	Ser	Ala	Pro	Leu	Ile
						260		265			270				

Thr Ala Met Cys Val Leu Ala Asn Ala Thr Phe Pro Cys Phe Gln Pro
 275 280 285

Pro Cys Val Pro Cys Cys Tyr Glu Asn Asn Ala Glu Ala Thr Leu Arg
 290 295 300

Met Leu Glu Asp Asn Val Asp Arg Pro Gly Tyr Tyr Asp Leu Leu Gln
 305 310 315 320

Ala Ala Leu Thr Cys Arg Asn Gly Thr Arg His Arg Arg Ser Val Ser
 325 330 335

Gln His Phe Asn Val Tyr Lys Ala Thr Arg Pro Tyr Ile Ala Tyr Cys
 340 345 350

Ala Asp Cys Gly Ala Gly His Ser Cys His Ser Pro Val Ala Ile Glu
 355 360 365

~~A~~ Val Arg Ser Glu Ala Thr Asp Gly Met Leu Lys Ile Gln Phe Ser
 370 375 380

~~A~~ Gln Ile Gly Ile Asp Lys Ser Asp Asn His Asp Tyr Thr Lys Ile
 385 390 395 400

~~A~~ Tyr Ala Asp Gly His Ala Ile Glu Asn Ala Val Arg Ser Ser Leu
 405 410 415

~~L~~ys Val Ala Thr Ser Gly Asp Cys Phe Val His Gly Thr Met Gly His
 420 425 430

~~P~~he Ile Leu Ala Lys Cys Pro Pro Gly Glu Phe Leu Gln Val Ser Ile
 435 440 445

~~G~~ln Asp Thr Arg Asn Ala Val Arg Ala Cys Arg Ile Gln Tyr His His
 450 455 460

Asp Pro Gln Pro Val Gly Arg Glu Lys Phe Thr Ile Arg Pro His Tyr
 465 470 475 480

Gly Lys Glu Ile Pro Cys Thr Thr Tyr Gln Gln Thr Thr Ala Lys Thr
 485 490 495

Val Glu Glu Ile Asp Met His Met Pro Pro Asp Thr Pro Asp Arg Thr
 500 505 510

Leu Leu Ser Gln Gln Ser Gly Asn Val Lys Ile Thr Val Gly Gly Lys
 515 520 525

Lys Val Lys Tyr Asn Cys Thr Cys Gly Thr Gly Asn Val Gly Thr Thr
 530 535 540

Asn Ser Asp Met Thr Ile Asn Thr Cys Leu Ile Glu Gln Cys His Val
 545 550 555 560

Ser Val Thr Asp His Lys Lys Trp Gln Phe Asn Ser Pro Phe Val Pro
 565 570 575

Arg Ala Asp Glu Pro Ala Arg Lys Gly Lys Val His Ile Pro Phe Pro
 580 585 590

Leu Asp Asn Ile Thr Cys Arg Val Pro Met Ala Arg Glu Pro Thr Val
 595 600 605

Ile His Gly Lys Arg Glu Val Thr Leu His Leu His Pro Asp His Pro
 610 615 620

Thr Leu Phe Ser Tyr Arg Thr Leu Gly Glu Asp Pro Gln Tyr His Glu
 625 630 635 640

Glu Trp Val Thr Ala Ala Val Glu Arg Thr Ile Pro Val Pro Val Asp
 645 650 655

Gly Met Glu Tyr His Trp Gly Asn Asn Asp Pro Val Arg Leu Trp Ser
 660 665 670

Gln Leu Thr Thr Glu Gly Lys Pro His Gly Trp Pro His Gln Ile Val
 675 680 685

Gln Tyr Tyr Tyr Gly Leu Tyr Pro Ala Ala Thr Val Ser Ala Val Val
 690 695 700

Gly Met Ser Leu Leu Ala Leu Ile Ser Ile Phe Ala Ser Cys Tyr Met
 705 710 715 720

Leu Val Ala Ala Arg Ser Lys Cys Leu Thr Pro Tyr Ala Leu Thr Pro
 725 730 735

Gly Ala Ala Val Pro Trp Thr Leu Gly Ile Leu Cys Cys Ala Pro Arg
 740 745 750

Ala His Ala Ala Ser Val Ala Glu Thr Met Ala Tyr Leu Trp Asp Gln
 755 760 765

Asn Gln Ala Leu Phe Trp Leu Glu Phe Ala Ala Pro Val Ala Cys Ile
 770 775 780

Leu Ile Ile Thr Tyr Cys Leu Arg Asn Val Leu Cys Cys Cys Lys Ser
 785 790 795 800

Leu Ser Phe Leu Val Leu Leu Ser Leu Gly Ala Thr Ala Arg Ala Tyr
 805 810 815

Glu His Ser Thr Val Met Pro Asn Val Val Gly Phe Pro Tyr Lys Ala
 820 825 830

His Ile Glu Arg Pro Gly Tyr Ser Pro Leu Thr Leu Gln Met Gln Val
 835 840 845

Val Glu Thr Ser Leu Glu Pro Thr Leu Asn Leu Glu Tyr Ile Thr Cys
 850 855 860

Glu Tyr Lys Thr Val Val Pro Ser Pro Tyr Val Lys Cys Cys Gly Ala
 865 870 875 880

Ser Glu Cys Ser Thr Lys Glu Lys Pro Asp Tyr Gln Cys Lys Val Tyr
 885 890 895

Thr Gly Val Tyr Pro Phe Met Trp Gly Gly Ala Tyr Cys Phe Cys Asp
 900 905 910

~~Ser~~ Glu Asn Thr Gln Leu Ser Glu Ala Tyr Val Asp Arg Ser Asp Val
 915 920 925

~~Cys~~ Arg His Asp His Ala Ser Ala Tyr Lys Ala His Thr Ala Ser Leu
 930 935 940

~~Lys~~ Ala Lys Val Arg Val Met Tyr Gly Asn Val Asn Gln Thr Val Asp
 945 950 955 960

~~Val~~ Tyr Val Asn Gly Asp His Ala Val Thr Ile Gly Gly Thr Gln Phe
 965 970 975

~~Ile~~ Phe Gly Pro Leu Ser Ser Ala Trp Thr Pro Phe Asp Asn Lys Ile
 980 985 990

Val Val Tyr Lys Asp Glu Val Phe Asn Gln Asp Phe Pro Pro Tyr Gly
 995 1000 1005

Ser Gly Gln Pro Gly Arg Phe Gly Asp Ile Gln Ser Arg Thr Val Glu
 1010 1015 1020

Ser Asn Asp Leu Tyr Ala Asn Thr Ala Leu Lys Leu Ala Arg Pro Ser
 1025 1030 1035 1040

Pro Gly Met Val His Val Pro Tyr Thr Gln Thr Pro Ser Gly Phe Lys
 1045 1050 1055

Tyr Trp Leu Lys Glu Lys Gly Thr Ala Leu Asn Thr Lys Ala Pro Phe
 1060 1065 1070

Gly Cys Gln Ile Lys Thr Asn Pro Val Arg Ala Met Asn Cys Ala Val
 1075 1080 1085

Gly Asn Ile Pro Val Ser Met Asn Leu Pro Asp Ser Ala Phe Thr Arg
 1090 1095 1100
 Ile Val Glu Ala Pro Thr Ile Ile Asp Leu Thr Cys Thr Val Ala Thr
 1105 1110 1115 1120
 Cys Thr His Ser Ser Asp Phe Gly Gly Val Leu Thr Leu Thr Tyr Lys
 1125 1130 1135
 Thr Asn Lys Asn Gly Asp Cys Ser Val His Ser His Ser Asn Val Ala
 1140 1145 1150
 Thr Leu Gln Glu Ala Thr Ala Lys Val Lys Thr Ala Gly Lys Val Thr
 1155 1160 1165
 Leu His Phe Ser Thr Ala Ser Ala Ser Pro Ser Phe Val Val Ser Leu
 1170 1175 1180
 Cys Ser Ala Arg Ala Thr Cys Ser Ala Ser Cys Glu Pro Pro Lys Asp
 1185 1190 1195 1200
 His Ile Val Pro Tyr Ala Ala Ser His Ser Asn Val Val Phe Pro Asp
 1205 1210 1215
 Met Ser Gly Thr Ala Leu Ser Trp Val Gln Lys Ile Ser Gly Gly Leu
 1220 1225 1230
 Gly Ala Phe Ala Ile Gly Ala Ile Leu Val Leu Val Val Val Thr Cys
 1235 1240 1245
 Le Gly Leu Arg Arg
 1250

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 115 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..115
- (D) OTHER INFORMATION: /label= 26S_region
/note= "26S promoter and transcription start and proximal downstream region of pSFV1; Figure 8."

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..24
- (D) OTHER INFORMATION: /product= "26S promoter region"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ACCTCTACGG CGGTCTAGA TTGGTGCCTT AATACACAGA ATCTGATTGG ATCCCGGGTA	60
--	----

ATTAATTGAA TTACATCCCT ACGCAAACGT TTTACGGCCG CCGGTGGCGC CCGCG	115
--	-----

(x) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 127 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..127
- (D) OTHER INFORMATION: /label= 26S_region
/note= "26S promoter and transcription start and proximal downstream region of pSFV2; Figure 8."

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..24
- (D) OTHER INFORMATION: /product= "26S promoter region"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACCTCTACGG CGGTCTAGA TTGGTGCCTT AATACACAGA ATTCTGATTA TAGCGCACTA	60
--	----

TTATATAGCA CCGGATCCCG GGTAAATTAT TGACGCAAAC GTTTACGGC CGCCGGTGGC	120
--	-----

GCCCGCG

127

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..123
- (D) OTHER INFORMATION: /label= 26S_region
/note= "26S promoter and transcription start and proximal downstream region of pSFV3; Figure 8."

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..24
- (D) OTHER INFORMATION: /product= "26S promoter region"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ACCTCTACGG CGGTCCCTAGA TTGGTGCGTT AATACACAGA ATTCTGATTA TAGCGCACTA 60

TTATATAGCA CCATGGATCC CGGGTAATTAA ATTGACGTTT TACGGCCGCC GGTGGCGCCC 120

GCG 123

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Semliki Forest Virus

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..54
- (D) OTHER INFORMATION: /label= restrict_site
/NOTE= "sequence of SFV E2 genome in vicinity of Bam HI site
in SFV vector E2; Figure 12."

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..54

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AAC	TCA	CCT	TTC	GTC	CCG	AGA	GCC	GAC	GAA	CCG	GCT	AGA	AAA	GGC	AAA		48
Asn	Ser	Pro	Phe	Val	Pro	Arg	Ala	Asp	Glu	Pro	Ala	Arg	Lys	Gly	Lys		
1				5						10					15		

GTC	CAT																54
Val	His																

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Asn	Ser	Pro	Phe	Val	Pro	Arg	Ala	Asp	Glu	Pro	Ala	Arg	Lys	Gly	Lys		48
1				5						10					15		

Val His

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: HIV

(ix) FEATURE:
 (A) NAME/KEY: -
 (B) LOCATION: 1..46
 (D) OTHER INFORMATION: /label= fragment
 /note= "HIV gp120 epitope introduced into SFV
 vector E2; Figure 12."

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1..45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAT	CCG	CGT	ATC	CAG	AGA	GGA	CCA	GGA	AGA	GCA	TTT	GTT	GAG	CTA		45
Asp	Pro	Arg	Ile	Gln	Arg	Gly	Pro	Gly	Arg	Ala	Phe	Val	Glu	Leu		
							5				10			15		

G	T	E	R	D	O	N	O		46
---	---	---	---	---	---	---	---	--	----

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asp	Pro	Arg	Ile	Gln	Arg	Gly	Pro	Gly	Arg	Ala	Phe	Val	Glu	Leu		
							1				5			10		15

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..51
- (D) OTHER INFORMATION: /label= chimaeric_seq
 /note= "SFV-HIV chimaeric sequence shown in Figure 12."

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..51
- (D) OTHER INFORMATION: /product= "SFV-HIV chimaeric sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAG GAT CCG CGT ATC CAG AGA GGA CCA GGA AGA GCA TTT GTT GAG GAT	48
Glu Asp Pro Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Glu Asp	
1 5 10 15	

CCG	51
Pro	

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Glu Asp Pro Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Glu Asp	
1 5 10 15	

Pro

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..60
- (D) OTHER INFORMATION: /label= oligonucleotide
/note= "used to introduce new linker site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGCCAGTGA ATTCTGATTG GATCCCGGGT AATTAATTGA ATTACATCCC TACGCAAACG

60

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..62
- (D) OTHER INFORMATION: /label= oligonucleotide
/note= "used to introduce new linker site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCGCACTATT ATAGCACCGG CTCCCGGGTA ATTAATTGAC GCAAACGTTT TACGGCCGCC

60

GG

62

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..62

(D) OTHER INFORMATION: /label= oligonucleotide
 /note= "used to introduce new linker site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCGCACTATT ATAGCACCAT GGATCCGGGT AATTAATTGA CGTTTACGG CCGCCGGTGG

60

SEQUENCE CHARACTERISTICS

62

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..21

(D) OTHER INFORMATION: /label= primer
 /note= "SP1 upstream sequencing primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGCGGTCTT AGATTGGTGC G

21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..21

(D) OTHER INFORMATION: /label= primer
 /note= "SP2 downstream sequencing primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CGGGCGGCC ACCGGCGGCC G

21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..21

(D) OTHER INFORMATION: /label= primer
 /note= "primer-1 for first strand cDNA synthesis"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTTCTCGTAG TTCTCCTCGT C

21

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..27

(D) OTHER INFORMATION: /label= primer
 /note= "primer-2 for first strand cDNA synthesis"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GATATCCAG TGGTTGTTCT CGTAATA

27

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..28

(D) OTHER INFORMATION: /label= primer
 /note= "5' most primer for second strand cDNA
 synthesis, equals bp 1-28 of SFV sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGGCGGATG TGTGACATAC ACGACGCC

28

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs
 (B) TYPE: nucleic acid

- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..46
- (D) OTHER INFORMATION: /label= adaptor
 /note= "5'-sticky end
 (EcoRI-HindIII-NotI-XbaIII-SpeI) blunt end-3'
 adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AATTCAAGCT TGCAGCCGCA CTAGTGTTCG AACGCCGGCG TGATCA

46

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..8
- (D) OTHER INFORMATION: /label= oligonucleotide
 /note= "NcoI oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCCATGGC

8

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..20
- (D) OTHER INFORMATION: /label= oligonucleotide
 /note= "oligonucleotide used for screening by colony hybridization"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GATGACACTA TAGCCATGGC

20

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..24
- (D) OTHER INFORMATION: /label= oligonucleotide
 /note= "site-directed mutagenic oligonucleotide used to introduce a BamHI site into the SFV genome"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GATCGGCCTA GGAGCCGAGA GCCC

24

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Semliki Forest Virus

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..80
- (D) OTHER INFORMATION: /label= terminator
 /note= "3' terminal sequence of cDNA expression
 vector complementary to alphavirus genomic RNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
TTCCAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
AAAAAAAAAA AAAACTAGT
```

60

80

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Semliki Forest Virus

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..54
- (D) OTHER INFORMATION: /label= `restrict_site`
`/note= "sequence of SFV vector E2 in vicinity of Bam HI site;`
`Figure 12."`

(ix) FEATURE:

- (A) NAME/KEY: mutation
- (B) LOCATION: 27..32
- (D) OTHER INFORMATION: /label= `restriction_site`
`/note= "BamHI recognition sequence introduced into`
`SFV E2 genome in SFV vector E2."`

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..54

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAC	TCA	CCT	TTC	GTC	CCG	AGA	GCC	GAG	GAT	CCG	GCT	AGA	AAA	GGC	AAA	48
Asn	Ser	Pro	Phe	Val	Pro	Arg	Ala	Glu	Asp	Pro	Ala	Arg	Lys	Gly	Lys	
1					5					10					15	

G	G	C	T	A	T
Val	His				

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Asn	Ser	Pro	Phe	Val	Pro	Arg	Ala	Glu	Asp	Pro	Ala	Arg	Lys	Gly	Lys	48
1					5					10					15	

G	G	C	T	A	T
Val	His				